IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 11, No. 1, March 2022, pp. 65~80
ISSN: 2252-8938, DOI: 10.11591/ijai.v11.i1.pp65-80

A deep learning-based multimodal biometric system using score fusion

Chahreddine Medjahed[1], Abdellatif Rahmoun[2], Christophe Charrier[3], Freha Mezzoudj[4]
[1]EEDIS Laboratory, Department of Computer Science, University of Djillali Liabes, Sidi Bel-Abbes, Algeria
[2]Department of Computer Science, ESI-SBA, University of Djillali Liabes, Sidi Bel-Abbes, Algeria
[3]GREYC Laboratory, Department of Multimedia and Internet, University of Caen Normandie, Caen, France
[4]Department of Computer Science, University Hassiba Benbouali of Chlef, Chlef, Algeria

Article Info

Article history:

Received Jul 8, 2021
Revised Nov 27, 2021
Accepted Dec 10, 2021

Keywords:

Biometric identification system
Deep learning
Identication
Multimodal biometric system

ABSTRACT

Recent trends in artificial intelligence tools-based biometrics have overwhelming attention to security matters. The hybrid approaches are motivated by the fact that they combine mutual strengths and they overcome their limitations. Such approaches are being applied to the fields of biomedical engineering. A biometric system uses behavioural or physiological characteristics to identify an individual. The fusion of two or more of these biometric unique characteristics contributes to improving the security and overcomes the drawbacks of unimodal biometric-based security systems. This work proposes efficent multimodal biometric systems based on matching score concatenation fusion of face, left and right palm prints. Multimodal biometric identification systems using convolutional neural networks (CNN) and k-nearest neighbors (KNN) are proposed and trained to recognize and identify individuals using multi-modal biometrics scores. Some popular biometrics benchmarks such as FEI face dataset and IITD palm print database are used as raw data to train the biometric systems to design a strong and secure verification/identification system. Experiments are performed on noisy datasets to evaluate the performance of the proposed model in extreme scenarios. Computer simulation results show that the CNN and KNN multi-modal biometric system outperforms most of the most popular up to date biometric verification techniques.

Corresponding Author:

Chahreddine Medjahed
EEDIS Laboratory, Department of Computer Science, University of Djillali Liabes
Sidi Bel-Abbes, 22000, Algeria
Email: chahreddine.medjahed@univ-sba.dz

1. INTRODUCTION

With the growth of intelligence artificial systems and e-technologies nowadays, personal biometric authentication has become an essential demanded technic: widely used in airports, buildings, mobile phones, identity cards and so on. The use of biometrics data is essential for learning powerful recognition systems. Many physiological traits (such as face, iris, fingerprint, palm-print, hand geometry, ear.) or behavioural ones (such as gait, signature, voice) are used to identify a person. These characteristics will not be lost or forgotten and can be used to distinguish one individual from another. The fusion of two or more of these characteristics contributes to improving the security and showing high performance and remedying the limits and the disadvantages of the unimodal biometric systems.

Face detection [1] task has the goal to detect all the human's faces in an image or sequence of images. Also, face identification (or recognition) [2], [3] system have the goal to detect a face in an image

and then using classifiers or matching algorithms to identify or recognize who the face belongs to. However, face analysis [4] is the technique to examine an image and extract information, such as age, sex, complexion, emotion and so on.

Face identification is useful in variety of daily life areas such as healthcare system, authentification operation and so on. Face recognition is a convenient technic because it is easy to collect faces data without active cooperation of the person and faces data are more representative and discriminant for recognition. However, other biometric features, can be used to recognize individual such as palm print, fingerprint, gaits, signature, speech and so on.

Recently, palm print [5] has become one of the most notable biometric recognition systems and it has received interest of researchers. Many advantages led to use this trait such as less distortion, rich features and high accuracy. The principal lines, ridges and wrinkles in structure of palm print are stable all through the life of a person.

In general, there exist two types of the biometric systems: unimodal biometric systems and multimodal ones. Fisrstly, for unimodal biometric systems, one trait is used to identify a person. These systems can encounter different degradations and limitations such as lack of distinctiveness of the biometric trait or nonuniversality, noisy sensor data and so on. As a solution to these kinds of problems, multimodal biometric systems are created using many biometrics traits. This fusion reduces the risk of any spoofing or faking of other identities.

According to the literature, for multimodal system, the different traits are fused at one of these levels: data-sensor, feature-extraction, matching-score and decision levels. Recently, researchers are more interested by the fusion at matching-score level because of its better recognition accuracy compared to the other levels. According to [6], "the score level fusion is commonly preferred in multimodal biometric systems because matching scores contain sufficient information to distinguish between genuine and impostor cases".

This paper introduces and compares many unimodal and multimodal biometric systems for human identification. The authors present strong multimodal biometric systems with deep learned and fuzzed scores' of three traits: face, left and right palm prints. First, the score of each modality is obtained using convolutional neural network (CNN) then, the fusion of scores helps to perform fusion at this level. The fusion of these modalities is implemented on Score Level using concatenation strategy. Second, k-nearest neighbors (KNN), the machine learning algorithm which remains a strong and a successful algorithm [7], [8] is used for the classification step. For more accurate evaluations and challenging situations, different kinds of biometric data are used: clear and noisy ones. Some variations in rotation and adding noises introduce large changes in faces and palms' images.

The rest of this paper is organized: an overview of previous works about multimodal biometric systems is presented in section 2. Section 3 summerises the techniques of deep learning neural networks used for scores learning and some machine learning tools dedicated for the classification. Section 4 describes the methodology of the used approach. Section 5 explores the experimental results. Section 6 concludes the work conducted and proposes some future works.

2. RELATED WORKS

Several works have demonstrated that a multi-modal biometric system can surpass some of inconveniencs of unimodal biometric system [9]. Many studies have suggested that by using information from multiple biometric traits, better performance can be achieved. In [10], Ross and Govindarajan have proposed multimodal biometric systems based on fusion of face and hand at feature level. Three different scenarios were developed. Firstly, a fusion of principal components analysis (PCA) and linear discriminant analysis (LDA), the principal components analysis and linear discriminant analysis algorithms respectively, face's coefficients was used. Second, a fusion of LDA coefficients which respesent the three channels of the face image: the red, green and blue was used. Finaly, fusion of face and hand traits was presented.

In [11], the authors proposed a fusion technique based on a discrete cosine transform (DCT) algorithm. A fused feature vector of face and palmprint data was constructed. The identification is done using gaussian mixture models (noted gaussian mixture model (GMM)). The proposed method produces good recognition rates when evaluated on FERET-PolyU and ORL-PolyU databases.

In [12], multimodal biometric was implemented based on fusion of retina, fingerprint and finger vein at feature level. The techniques such as blood vessel extraction, minutia extraction and maximum curvature were used to extract the useful features. The fuzed features are encrypted using the asymmetric public-key cryptosystem algorithm of rivest shamir adleman (RSA) and compared to a stored template to authenticate the person. The use of the RSA algorithm improves the baseline multimodal biometric's performance.

In [13], multimodal biometric systems with fusing the face, the palm print at different levels, sensor level, feature level, score level and decision level were introduced. The proposed systems were evaluated on the available publically PolyU and AR datasets for the palm print and face respectively. The result of this search showed the best perfomance is obtained with the score level fusion using sum rule with an accuracy of 97.5%.

In [14], the authors introduced a Multimodal biometric recognition system by combining face and both left iris and right iris. For face trait, the features were extracted with deep belief network (DBN). By applying CNN for each trait, the scores obtained were fused at two different levels: rank level and score level. Many databases were used to realize this work such as the facial recognition technology (FERET) database, SDUMLA HMT and CASIA V1.0.

In [15], the authors proposed multi-biometric systems for human verification using CNN to fuse iris and face traits on feature and score levels. They utilized the very deep CNN called VGG16 [16] to extract features from images. The recognition step is based only on the features without using any image detection techniques. The experimentations were conducted on the multimodal biometric database SDUMLA-HMT.

In our case, the main objective is to evaluate the performance of unimodal and multimodal biometric systems. As multimodalities, we use the fusion of the face, the right palm print and the left palm print traits at score level. The proposed models are based on deep learning models for feature extraction and machine learning tools for classification task, as illustrated in Figure 1. The evaluations of the proposed approach are done using clear and noisy and rotated data.

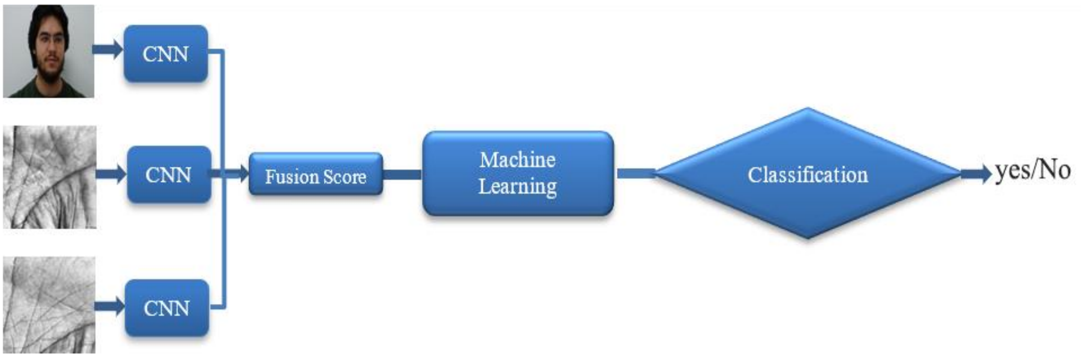

Figure 1. Pipeline tasks of our proposed methodology

3. MACHINE LEARNING APPROACHES

The main useful tools for our biometric systems are described. Three steps are involved is: (i) data pre-processing, (ii) feature extraction using the deep learning algorithms and (iii) training and testing identification person models using Machine learning algorithms.

3.1. Pre-processing

Differences in aging, occlusion, facial expressions, noises and poses faces' images constitute complex challenges for face recognition systems. In general, it is crucial, before any biometric recognition, to apply the face alignment, which contribute to detect the face area and to remove the background. Also, many technics are considered as image pre-processing and are used to enhance the quality of the data and facilitate the recognition task such as alignment face (or palm print), normalisation and de-noising.

However, other types of technics, such as deformation, scaling, rotation, changing colors, adding noises and so on, are applied on the original images for the data augmentation. In our case, we use some of these technics, such as adding noises and applying rotation, to decrease the images quality. Our goal is to obtain more challenging data as we can find in difficult or critical real situations.

The external disturbance such as environmental conditions during data acquisition or the quality of the sensing elements themselves can cause noise [17]. In this paper, we explore two types of noises: salt-and-pepper noise and gaussian noise. Also, rotations with different degrees are applied to the intial used data.

3.2. Convolutional neural networks

CNN are popular tools in the field of deep learning. Their robustness is due to their flexible architecture and their ability to extract features from raw data. They are successfully used in image classification [18], objects detection, Speech recognition and language modeling [19].

To accelerate the modeling and avoid the expensive computation and decrease the over-fitting due to the lack of the labeled data in some fields, many studies tuned and used deep pre-trained models (e.g,, AlexNet [20], VGG [16], GoogleNet [21], Resnet [22] and so on), as shown in Table 1. For image recognition task, the CNN input is an image with red, green and blue (RGB) channels and the output is the prediction of the image's category. These CNNs, mentioned above, are pre-trained on the dataset ImageNet [23] which is a dataset for computer vision research with more than 14 million of images.

In general, to train and test a CNN, series of convolution, pooling and fully connected layers are applied, followed by Softmax function to classify the data. These operations are the basic building blocks of every CNN. The kernel trick help to transform nonlinear case to linear one.The kernel size is choosen according to the variation in the lacation of the input information [24]. The *inception* [21] technic helps to have filters with multiple sizes operating on the same level**.** The *Dropout* [25] is used for the neural network regularization, which helps to reducing interdependent learning amongst the neurons.

Table 1. Comparison of some pre-trained CNNs

	AlexNet [20]	VGGNet [16]	GoogleNet [21]	RestNet [22]
#layers (convolutional + fully connected)	5+3	13/16+3	21+1	151+1
Kernel size	11, 5, 3	3	7, 1, 3, 5	7, 1, 3, 5
Data Aug.	+	+	+	+
Inception [21]	-	-	+	-
Dropout [25]	+	+	+	+

3.2.1. Convolutional layer

The convolutional layer is used to extract discriminative features from the images. "This bloc contains a set of convolution kernels (called filters). They are convolved with the input and generate a "feature map" [25]. Mathematically, the convolution procedure can be expressed using the (1):

$$y_{in} = f(\sum_{j=0}^{n} w_j\, x_j) \tag{1}$$

where x is an input value from the image, w is the weight value from the filter, the pixel number is noted by j. The function f is an activation function. The rectified activation function (ReLU) is widely used in Deep Learning. It replaces negative values with zero, according to the (2) as shown in Figure 2, where z is the convolutional layer output [26]:

$$f(u) = max(\,0, u) \tag{2}$$

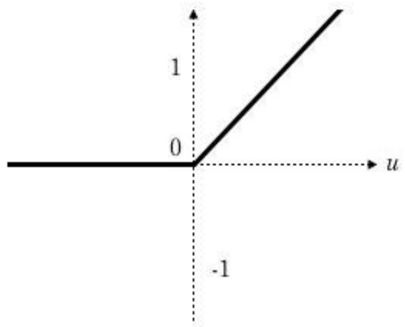

Figure 2. The ReLU function

3.2.2. Pooling layer

The second operation after convolution in the CNN is the Pooling. The pooling operation is helpful for acquiring a reduced component portrayal, which is invariant to direct changes in object scale, pose, and translation in an image [25]. Two kind of pooling operation are widely used: max pooling and average pooling. Max pooling compute the maximum element of the selection. It is most used type because it is fast to calculate and allows to effectively simplifying the image. For the average pooling, we calculate the average of the selection as shown in Figure 3.

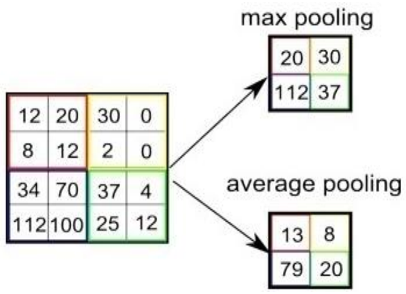

Figure 3. Pooling operations

3.2.3. Fully connected layer

The fully-connected layers are in the CNN's top. To facilite the classification task, it is necessary to convert the outputs of these previous fully-connected layers to probabilities. The *softmax function* is able to calculate them using the (3), where *m* is the class, *n* is the maximum classes' number, the output *y* is computed using the (4), where *x* is the feature vector of the data sample and *w* respresent the weight vector. The softmax classifier output which is a score vector represent a set of probabilities according to the different classes [26]:

$$soft\,max(i) = \frac{e^{y_i}}{\sum_{m=1}^{n} e^{y_m}} \tag{3}$$

$$y_{out} = \sum_{l=0}^{k} w_l\, x_l \tag{4}$$

3.3. Training convolution neural network

For training the CNN, "a loss function is used to estimate the quality of the prediction. This function quantifies the difference between the prediction made by the model and the correct output [25]. Training CNN is finding the best parameters of the network to reduce this function. There exist many types of loss function, such as: mean squared error, cross entropy loss and hinge loss. The type fuction must be choosen according to the traited problem. Gradient descent is the optimization algorithm employed to minimize the error by computing the gradient required for updating network parameter values.

3.3.1. AlexNet

AlexNet [20] is the first successful CNN for big data. It has a similar architecture to the original LeNet but it is deeper and wider CNN model. The architecture of AlexNet as shown in Figure 4 contains eight layers, five convolutions layer with max pooling and three fully connected layers. There are 60 million learning parameters and 650,000 neurons. AlexNet is the first CNN that uses ReLU activation function. The input of this CNN is RGB image with a size of 227×227×3.

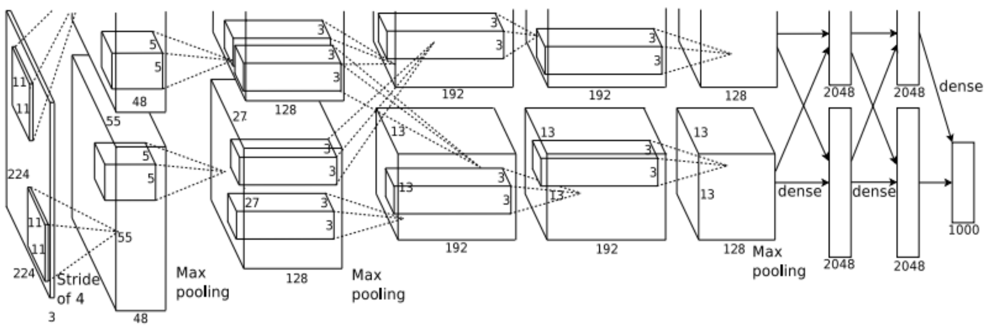

Figure 4. The architecture of AlexNet [20]

3.3.2. GoogleNet

In 2014, GoogleNet [21] has achieved the best result in ImageNet large scale visual recognition challenge (ILSVRC), the ImageNet large scale visual recognition challenge. Googlenet uses fewer

parameters than the CNN AlexNet. GoogleNet implements Inception modules with the aim of optimizing the usage of computing resources within the network. The idea is to apply parallel pooling and convolutions operations with different kernel sizes and to concatenate the resulting feature maps before going to the next layer. GoogleNet has in total 22 layers and it uses an average pooling. The input of this CNN is RGB image with a size of 224×224×3 as shown in Figure 5.

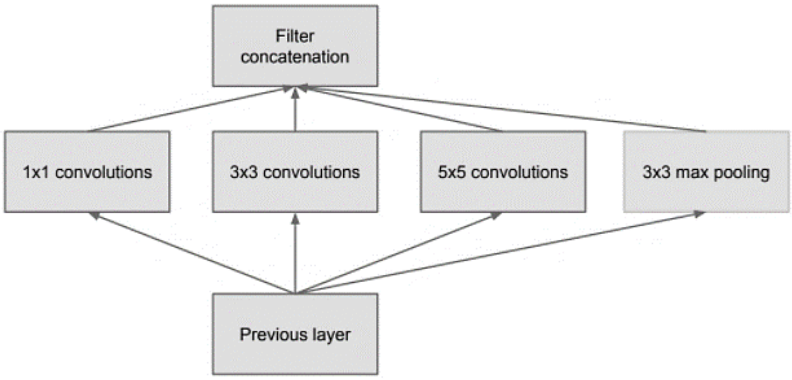

Figure 5. Simple scheme of an inception block as proposed by [21]

3.3.3. ResNet

In 2015, the Microsoft's residual network ResNet [22] has achieved the best result in ILSVRC, the ImageNet large scale visual recognition challenge. It was proposed with a residual learning block. Resnet overcome the problem of vanishing gradient and it is developed with different layers 18, 34, 50, and 101. The residual network architecture's remarkable feature is the identity skips connections within the residual blocks, which enables very deep CNN architectures to be trained easily. The residual network consists of several residual blocks which are stacked on top of each other [25], as illustrated in Figure 6.

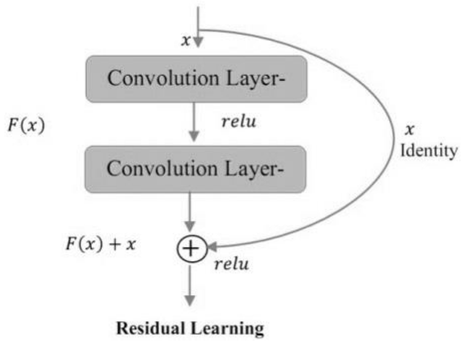

Figure 6. ResNet residual learning building block

3.4. Machine learning algorithms

Machine learning methods [27] are important tools for researchers, scientists and students in a wide range of areas. Traditionally, different techniques like k-nearest neighbors algorithm and support vector machines are used for the face recognition tasks [28]. These methods are based on hand crafted data representation such as detection of regions of interest and feature extraction. Among the feature's extraction methods: eigenfaces local and binary patterns. Are used. However, in our case, we use these machine learning methods to classify the scores obtained by the deep CNNs.

3.4.1. Naive Bayes

One straightforward source of classifier based on probability computation is the famous naïve Bayes classifier. There are many variants of this algorithm but all focus on the strong and naïve independence

assumption between the features. "The naïve Bayes assumption is helpful when the dimensionality D of the input space is high, making density estimation in the full D-dimensional space more challenging". This supervised learning algorithm uses the famous Bayes theorem [24], [27].

3.4.2. Support vector machines (SVM)

SVM work on induction principle, called structural risk minimization, which targets to minimize an upper bound on the expected generalization error. The SVM uses the concept of mathematical planes, called maximum-margin hyperplanes, to distinguish between the different classes. It draws a plane between two classes. The SVM training consists on trying to maximize the distance of this plane from both classes using the concept of support vectors, which are the outermost points of each class. This margin is drawn explicitly in the case of a linear classification" [28], [29]. Also, inorder to find the hyperplane, the SVM uses the kernel trick with nonlinear classifications to transform nonlinear case to linear one. Also, the SVM was, first formulated for binary classification, and the extension to multi-classes is useful [27].

3.4.3. AdaBoost

In 1996, Freund and Schapire have developed AdaBoost, for adaptive boosting, which is an algorithm for combining many simple weak classifiers to obtain a strong classifier using a linear combination. It is a popular algorithm of machine learning that has the advantages of being quick in term of speed, easy to be programmed, simple in operation and there is no need to adjust parameters except for the number of iterations. AdaBoost algorithm generates a collection of bad learners by maintaining a weight over training data and adjusting them to each "weak period". The weights of the training samples misclassified by current poor learners will be increased while the weight of the correctly identified samples will be reduced [27].

3.4.4. Subspace discriminant

Subspace discriminant [27], [30] has been abundantly studied in data mining and pattern recognition. It is often combined and improved by the LDA which provid low-dimension for the discriminant subspace. Many studies have been performed to investigate the impact on the effectiveness on classification success in the ensemble learning of different subspacing, weighting and resampling techniques. Subspace discriminant model uses a random subspace algorithm to construct an ensemble of discriminant classifiers [25].

3.4.5. K-nearest neighbors

The supervised machine learning algorithm k-nearest neighbors (KNN) is based exclusively on the choice of classification metric. It is non-parametric, k must be fixed, and it is based on training data. The algorithm allows making a classification without making a hypothesis on the function $y = f(x_1, x_2, \ldots, x_p)$ which links the dependent variable to the independent variables.

The generalized distance between two variables is calculated using (5):

$$L_q = \left(\sum_{i,j=1}^{k} \left|x_i - x_{j\neq i}\right|^q\right)^{\frac{1}{q}} \tag{5}$$

when q=2, it is referred to euclidean distance and manhattan distance. The nearest neighbor is the variable with the shortest distance possible [7], [8], [27].

4. THE PROPOSED APPROACH DEEP LEARNING-BASED MULTIMODAL BIOMETRIC SYSTEM USING SCORE FUSION (DLMBS)

This section proposes a DLMBS. Firstly, we must identify which type of CNN is the best fit for such types of biometric data: face, left palm and left palm. These will be trained separately (or eventually simultaneously depending on the type of machine) up to feature layers at the score level (feature vectors). Then, score vectors will be fused to construct a multi-modal feature score. A separate experiment will be conducted to come up with the most performing way to combine such scores (linear combination, arithmetic averaging, and concatenation). This will be an input to a CNN that performs a final classification.

Other experiments will be done to test several machine learning (ML) classifiers. According to the best fit, we choose the best algorithm to construct the hybrid person identification system. The hybrid deep learning (DL), CNN, and ML models are based multi-modal scores. We notice that all these experiments are conducted using clean data.

With a similar scenario, we will test the effect of simulated noisy and oriented data on the proposed models. Two kinds of noises are introduced on the initial clean data. Also, some geometrical deformations are

applied on the clean data. These simulated challenging situations help to test the robustness of the DLMBS performance.

4.1. Preliminary experiment

We create three separate unimodal biometric systems, based on respectively face, left palm, and right palm. Each of these biometric systems uses different types of CNN; ie; Alexnet, Googlenet and Resnet-18 neural networks respectively. These are also trained separately using standard datasets: FEI Face Dataset [31] and IITD Palm print Database [32] as shown in Figure 7.

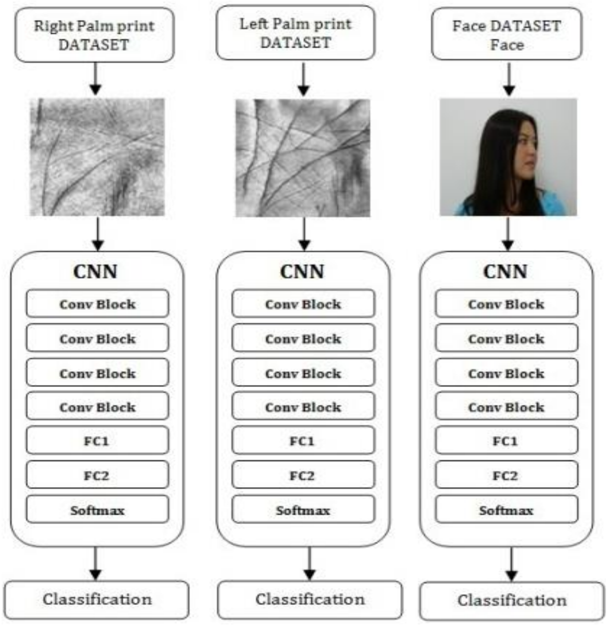

Figure 7. Part 1-unimodal biometric CNN models

4.1.1. Biometrics datasets

a) FEI Face dataset

The Brazilian FEI face database present a set of face images for 100 men and 100 women (200 individuals) that are students and staff of FEI laboratory between 19 and 40 years old. Each person has 14 images. Each image is with 640×480 pixels. All images are in color with different position of head, frontal pose and the head turning from left to right. Variations in illumination and head poses introduce large changes in images [31] as shown in Figure 8.

Figure 8. Sample of faces from FEI dataset with different head poses

b) IITD palm print V1 database

The IITD Palm print V1 database [32] as shown in Figure 9 is a hand database that contains a set of hand images with 800×600 pixels for 230 individuals that are students and staff at IIT Delhi campus, with 12-57 years old. Six or seven images from each subject, for each of the left and right palmprint, are acquired in different hand pose. Apart from the original images, there are also automatically cropped 150×150 pixels and normalized palm print images.

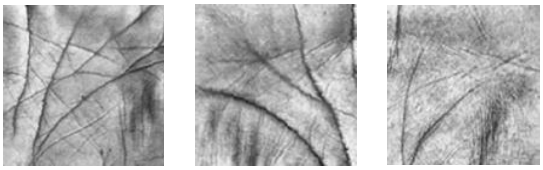

Figure 9. Sample of cropped Images from IITD Palm print V1 Dataset: each palm consists on principal lines, wrinkles and epidermal ridges

In this paper, we choose to use a subset that contains 140 subjects' faces from FEI face database and 140 hands' subjects from IITD palm print V1 database. Each subject has five different images from the three modalities (face, left palm print and right palm print) for training purposes. The training/testing ratio will be 80-20% respectively.

Thus, input matrices to CNNs will be of dimension (140×5) for each, and the values will be normalized to 1. We assign rows of the left palm prints and rows of the right palm prints to the corresponding rows of the face matrix. For sake of experimental convenience, we assume that every row (face, left palm, right palm) belongs to the same person, even though the two datasets FEI Face and IITD Palm print are of two different populations.

4.1.2. Training CNNs for separate modalities

In this section; we make an image classification for each modality; each modality is trained independently. All faces and palm print images are resized to 227×227, for AlexNet, 224×224; for GoogleNet and for Resnet 18. Table 2 shows the results for the unimodal identification biometric systems. We notice that the Resnet 18 neural network performs best for the face biometric system and the left and right palm print respectively. The resNet is successfully used in many fields and these results coincided with the literature [33], [34]. Apparently, ResNet 18 neural networks give the best accuracy rates for all three modalities.

Table 2. Results for unimodal systems of face and palm prints

Modalities	CNN	Time of Classification [s]	Accuracy Rate [%]
Face	AlexNet	10.26	99.28
	GoogleNet	25.29	97.14
	ResNet18	28.48	100
Left Palm Print	AlexNet	30.81	92.14
	GoogleNet	23.21	85.71
	ResNet18	30.10	95.00
Right Palm Print	AlexNet	3.96	87.14
	GoogleNet	15.26	86.43
	ResNet18	13.87	95.71

4.2. Training multi-modal biometric system (clean data)

The multimodal biometric system is evaluated by combining the face and the palmprints traits at score level. Preliminary experiments show that a concatenation as a fusion technique performs better than other types of combinations. The principal of the proposed person identification models is illustrated in the Figure 10. Here also, the two datasets FEI face and IITD Palm print V1 are used for the three CNNs training. The obtained scores are fused subsequently, and then classified with different types of ML classifiers. The Table 3 summarizes the most important evaluation results of the conducted experiments.

We observe that the fusion of two (or the three) biometric traits (face and Palm prints' scores using Resnet 18), as shown in Table 3, gives the best performance. The classification using Machine Learning algorithms such as SVM or naïve Bayes gives weakest results comparing to the results obtained by KNN, Adaboost and Subspace discriminant. Moreover, the central processing unit (CPU) processing time required by KNN for the classification step is very short. Furthermore, Resnet 18 neural network associated with KNN performs best for a multi-modal biometric system.

4.3. Training multi-modal biometric system (noisy data)

In this section, we will simulated and evaluated the effect of environment disturbance on the images during the acquisition process. The diversity of the angle during the acquisition of the image or the orientation of the capture devices or the low-quality sureveillance camera can affect the images quality.

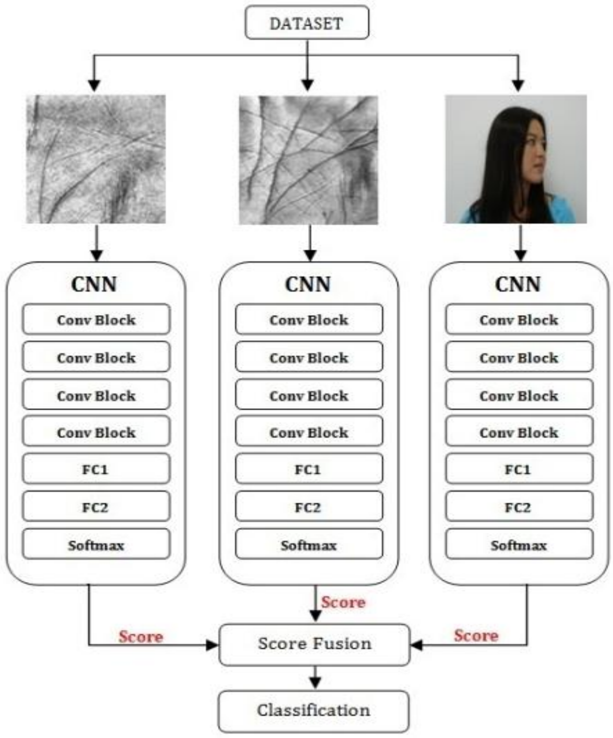

Figure 10. Multimodal biometric CNN model

Table 3. Results for multimodal systems of face, left and right palm prints (score level)

Modalities	Method	Time of Classification[s]	Accuracy Rate [%]
Face +	Naive Bayes	59.959397	92.86
Left Palm Print	SVM	1095.948599	78.57
	KNN	7.082602	100
	Adaboost	14.037205	100
	Subspace discriminant	44.535800	100
Face +	Naive Bayes	71.17050	89.28
Right Palm Print	SVM	1154.731083	77.86
	KNN	4.250831	100
	Adaboost	16.730207	100
	Subspace discriminant	59.260625	100
Left Palm Print +	Naive Bayes	54.982823	81.43
Right Palm Print	SVM	859.992817	57.14
	KNN	3.924279	100
	Adaboost	12.968442	93.57
	Subspace discriminant	41.010611	99.28
Face +	Naive Bayes	77.295227	92.14
Left Palm Print +	SVM	902.416240	83.57
Right Palm Print	KNN	5.248682	100
	Adaboost	13.430381	100
	Subspace discriminant	49.546649	100

4.3.1. The effect of noisy data on biometric systems

"Noise is a random variation of color information. It can affect the original signal and decrease it quality. Some external disturbances can be the cause such as: environmental conditions during image acquisition and the quality of the sensing elements themselves [17]. In order to simulate noisy data, we generate two kinds of noises: the Gaussian noise and the salt-and-pepper noise.

a) Salt and pepper noise

Salt-and-pepper noise in the images is due to faulty memory locations in hardware, malfunctioning pixels in camera and so on. The salt-and-pepper noise is also known as impulse noise, data drop noise or binary noise. Also, this type of noise can seen in the transmission of data and it appears as black dots on white background and white dots on a black one, as shown in Figure 11 [17].

Figure 11. Samples of face images with salt and pepper noise

b) Gaussian noise

Gaussian noise is also known as normal noise or white noise. Gaussian noise is caused by the discrete nature of warm object radiation and thermal atom vibration. [20]. The associated gaussian density function is given using the (6), also see the Figure 12:

$$P(z) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(z-\mu)^2}{2\sigma^2}} \tag{6}$$

where, the gray level is represented by z, the mean value is noted by μ, the standard deviation and the variation are noted by σ and σ^2 respectively.

Figure 12. Samples of face images with gaussian noise

Two different noises are added to the data, salt-and-pepper and gaussian noises; gradually for face, left palm print and right palm print separately, then we combined these traits in score level with different possible scenarios. Figure 13 shows clearly that face is more resistant to noise than palm prints for the salt-and-pepper noise. The similar results are obtained with the gaussian noise.

Also, we compared the multi-modal biometric system (fused scores) versus the models trained on the data with the both types of noises gaussian and salt and pepper. We use the three modalities in our experiments. According to the obtained results of the accuracy, it is clear that combining face, left and right palm prints give a very accurate verification biometric system. We conclude that CNN and KNN model is robust and isn't badly affected by noise.

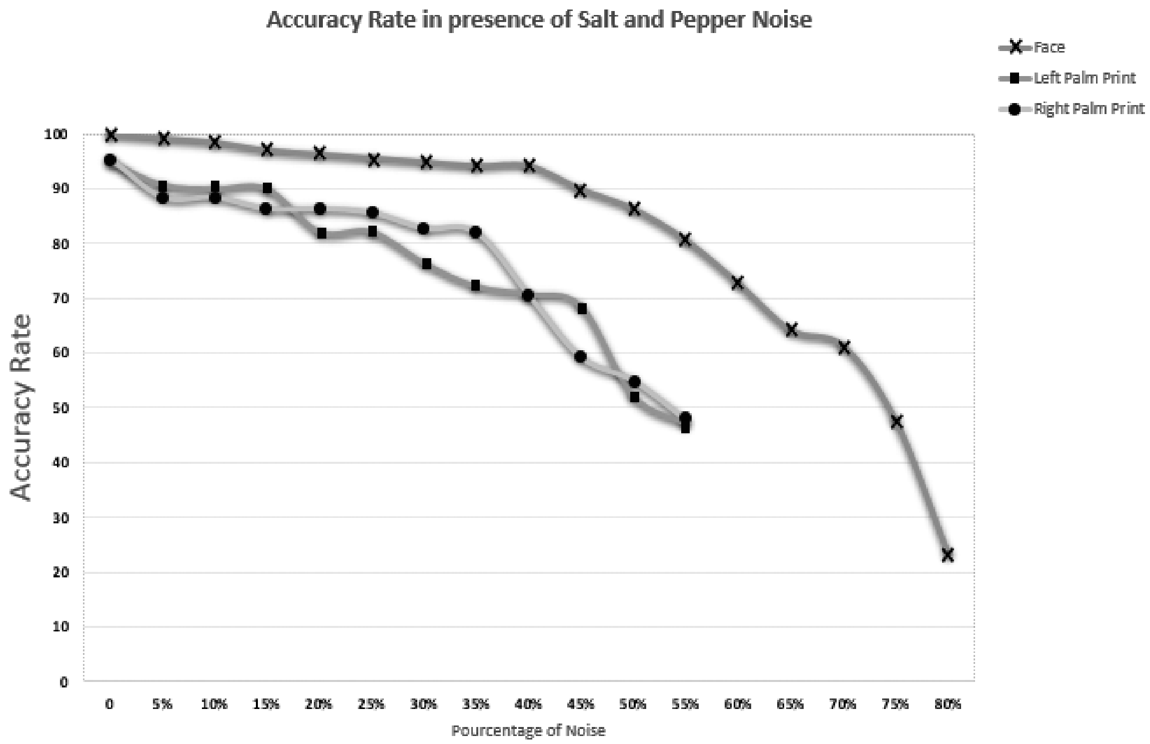

Figure 13. Accuracy rate in presence of salt and pepper noise

4.3.2. The effect of the geometrical deformation of images on biometric systems

We expose the experiments and their important results for image classification with different angle of rotation such as 0°, 30°, 45°, 60°, and 90°. We have generated multiple training images using rotation techniques from a training image. The principle is to use CNNs to analyse the classification performance on several variants of data as shown in Figure 14. The new simulated data involve both novel training and testing of the models. The Figure 15 shows verification accuracy for uni-modal systems using data with different degrees of rotation.

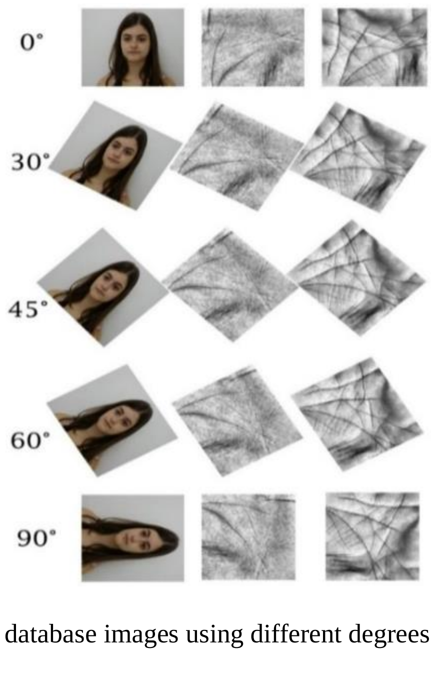

Figure 14. database images using different degrees of rotation

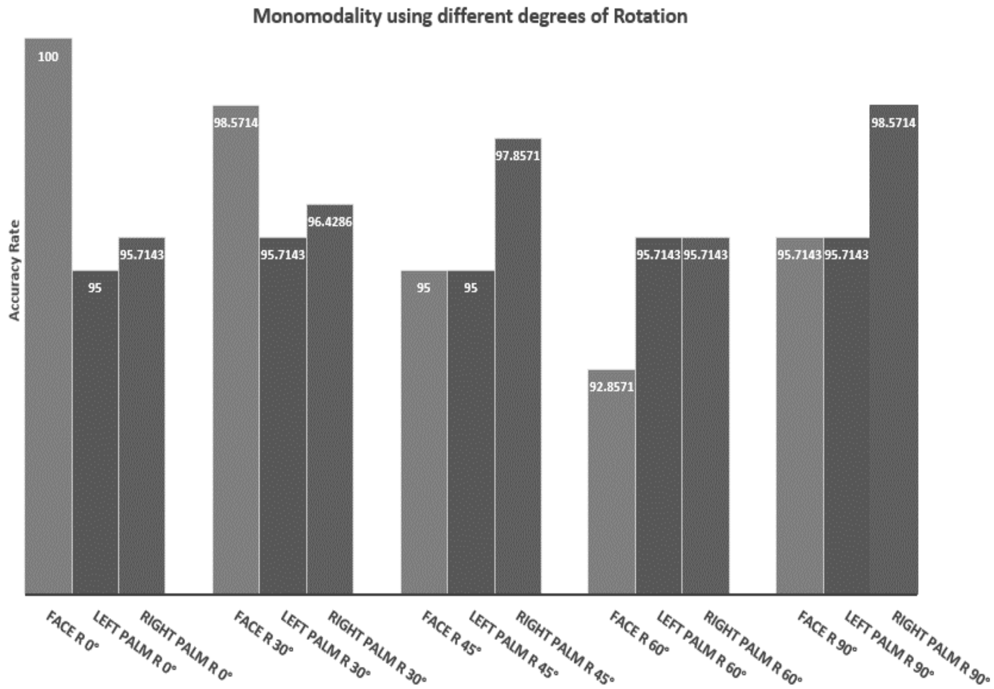

Figure 15. Accuracies rates for Monomodal systems using different degrees of rotation

We notice that the fusion of these traits contributes to decrease the performance of the biometric systems. The results obtained by the monomodal system based on face trait waved between 100% and 92.86% of accuracy rates, as it is shown in the Figure 15. However, the monomodal biometric systems, based on palm maintain performances between 98.57% and 92.86% for the best cases.

Many experiments are done with the rotated data. In general, the obtained results by the fusion of left and right palm prints with different angle of rotation respectively (e.g. 0°+30°, 0°+45%, and so on), achieve 100% of accuracy rates for all the situations. This phenomenon confirms our doubts about the fact that the rotation of these two traits (left and right palm prints) doesn't make sense for recognition and can conduct our biometric systems to over-fitting. And finally, the Figure 16 shows verification accuracy for multi--modal systems using different degrees of rotation.

With similar scenario, we fused the faces (with any rotation) and the left and the right palms. For example, for the two modalities face+(left|right) with a rotation of: 0°+30° and 0°+45°), the models achieve 100% of accuracy in all the situations. This phenomenon can be explained by the fact that the presence of the clean face image (without noise and without rotation) helps to enhance the performance of the multimodal biometric systems as much as possible.

In the Table 4, we present a comparison of our results with other recent works, which is not easy. The used databases, the data quality and the explored algorithms change and vary. However, we notice that our data are augmented and more challenging with adding the noise and the rotation. In addition, the recognition rate obtained with our system based on CNN and KNN is significantly good.

Table 4. Comparison of some recent works, including our system

Modalities	Databases Used	Rate Recognition	Reference
Face-Iris features level	FERET-CASIA v3.0	99.33%	[34]
ECG and Fingerprint decision level fusion feature and level fusion	-CYBHi database and PTB database -LivDet2015 fingerprint database and FVC 2004 database	Less than 100%	[35]
Face-Iris-Fingerprint (features level)	CMU, Multi-PIE, BioCop, and BIOMDATA	99.90%	[36]
Face-Palmprint (features level)	ORL-PolyU and FERET-PolyU	99.7%	[11]
Face-Palmprint (features level)	FERET face and PolyU palm print databases	99.17%	[37]
Face-Palmprint (left and right) Quality: (raw, clean, noisy) without and with: noise and rotation.	FEI face database, IITD Palm print V1	100%	Our System

Figure 16. Fusion left palm print and right palm print with different angle of rotation

5 CONCLUSION

In the present work, multimodal biometric identification systems are proposed using CNN and KNN. The fusion of modalities has proven the strengths of most biometric verification systems when it comes to security matters. The proposed model passed several steps during the design process to determine the best-fit CNN model, as well as the most significant classifier that can be suitable for three types of biometric modalities: face, left palm and right palm prints. The proposed model is then subjected to various types of noise and deformation added to the used data. The results of the conducted experimentations show clearly that the retained system is more resistant to such disturbance in terms of verification performance than any other unimodal biometric system. A de-noising pre-processing of the biometric data seems to be a good initiative to prevent verification performance degradation. The proposed method (CNN and KNN) can be used perfectly for clean and noisy data. Furthermore, future work will emphasize combining other biometrics data such as iris, voice, digital signature and handwriting. A larger-scale application domain such as government biometric data would use huge datasets, so it will be convenient to study the impact of dataset sizes on the performance of such systems. It will be interesting also to investigate other types of machine learning techniques to be associated with these biometric identification systems.

REFERENCES

[1] Y. Wu and Q. Ji, “Facial landmark detection: a literature survey,” *International Journal of Computer Vision*, vol. 127, no. 2, pp. 115–142, May 2018, doi: 10.1007/s11263-018-1097-z.
[2] D. S. Trigueros, L. Meng, and M. Hartnett, “Face recognition: from traditional to deep learning methods,” Oct. 2018, [Online]. Available: http://arxiv.org/abs/1811.00116.
[3] I. Masi, Y. Wu, T. Hassner, and P. Natarajan, “Deep face recognition: a survey,” in *2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, Oct. 2018, pp. 471–478, doi: 10.1109/SIBGRAPI.2018.00067.
[4] D. Poux, B. Allaert, J. Mennesson, N. Ihaddadene, I. M. Bilasco, and C. Djeraba, “Facial expressions analysis under occlusions based on specificities of facial motion propagation,” *Multimedia Tools and Applications*, vol. 80, no. 15, pp. 22405–22427, Jun.

2021, doi: 10.1007/s11042-020-08993-5.
[5] D. Zhong, X. Du, and K. Zhong, "Decade progress of palmprint recognition: A brief survey," *Neurocomputing*, vol. 328, pp. 16–28, Feb. 2019, doi: 10.1016/j.neucom.2018.03.081.
[6] M. He *et al.*, "Performance evaluation of score level fusion in multimodal biometric systems," *Pattern Recognition*, vol. 43, no. 5, pp. 1789–1800, May 2010, doi: 10.1016/j.patcog.2009.11.018.
[7] M. J. Alhasan, S. M. Abdulhussein, and A. H. K. Khwayyir, "Fingerprint positioning of users devices in long term evolution cellular network using k-nearest neighbour algorithm," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 1, pp. 528–535, Feb. 2021, doi: 10.11591/ijece.v11i1.pp528-535.
[8] E. G. Nihad, E.-N. El Mokhtar, Z. Abdelhamid, and A. A. Mohammed, "Hybrid approach of the fuzzy c-means and the k-nearest neighbors methods during the retrieve phase of dynamic case based reasoning for personalized follow-up of learners in real time," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9, no. 6, pp. 4939–4950, Dec. 2019, doi: 10.11591/ijece.v9i6.pp4939-4950.
[9] M. Singh, R. Singh, and A. Ross, "A comprehensive overview of biometric fusion," *Information Fusion*, vol. 52, pp. 187–205, Feb. 2019, doi: 10.1016/j.inffus.2018.12.003.
[10] A. A. Ross and R. Govindarajan, "Feature level fusion of hand and face biometrics," in *Biometric Technology for Human Identification II*, Mar. 2005, vol. 5779, doi: 10.1117/12.606093.
[11] M. I. Ahmad, W. L. Woo, and S. Dlay, "Non-stationary feature fusion of face and palmprint multimodal biometrics," *Neurocomputing*, vol. 177, pp. 49–61, Feb. 2016, doi: 10.1016/j.neucom.2015.11.003.
[12] D. Jagadiswary and D. Saraswady, "Biometric authentication using fused multimodal biometric," *Procedia Computer Science*, vol. 85, pp. 109–116, 2016, doi: 10.1016/j.procs.2016.05.187.
[13] T. A. Alghamdi, "Evaluation of multimodal biometrics at different levels of face and palm print fusion schemes," *Asian Journal of Applied Sciences*, vol. 9, no. 3, pp. 126–130, Jun. 2016, doi: 10.3923/ajaps.2016.126.130.
[14] A. S. Al-Waisy, R. Qahwaji, S. Ipson, and S. Al-Fahdawi, "A multimodal biometrie system for personal identification based on deep learning approaches," in *2017 Seventh International Conference on Emerging Security Technologies (EST)*, Sep. 2017, pp. 163–168, doi: 10.1109/EST.2017.8090417.
[15] N. Alay and H. H. Al-Baity, "A multimodal biometric system for personal verification based on different level fusion of iris and face traits," *Bioscience Biotechnology Research Communications*, vol. 12, no. 3, pp. 565–576, Sep. 2019, doi: 10.21786/bbrc/12.3/3.
[16] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," Sep. 2014, [Online]. Available: http://arxiv.org/abs/1409.1556.
[17] M. Koziarski and B. Cyganek, "Image recognition with deep neural networks in presence of noise - dealing with and taking advantage of distortions," *Integrated Computer-Aided Engineering*, vol. 24, no. 4, pp. 337–349, Sep. 2017, doi: 10.3233/ICA-170551.
[18] C. Medjahed, F. Mezzoudj, A. Rahmoun, and C. Charrier, "On an empirical study: face recognition using machine learning and deep learning techniques," in *Proceedings of the 10th International Conference on Information Systems and Technologies*, Jun. 2020, pp. 1–9, doi: 10.1145/3447568.3448521.
[19] F. Mezzoudj and A. Benyettou, "An empirical study of statistical language models: n-gram language models vs. neural network language models," *International Journal of Innovative Computing and Applications*, vol. 9, no. 4, 2018, doi: 10.1504/IJICA.2018.10016827.
[20] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.
[21] C. Szegedy *et al.*, "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2015, pp. 1–9, doi: 10.1109/CVPR.2015.7298594.
[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.
[23] J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, and Li Fei-Fei, "ImageNet: a large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2009, pp. 248–255, doi: 10.1109/CVPR.2009.5206848.
[24] E. Suryawati, R. Sustika, R. S. Yuwana, A. Subekti, and H. F. Pardede, "Deep structured convolutional neural network for tomato diseases detection," in *2018 International Conference on Advanced Computer Science and Information Systems, ICACSIS 2018*, Oct. 2019, pp. 385–390, doi: 10.1109/ICACSIS.2018.8618169.
[25] S. Khan, H. Rahmani, S. A. A. Shah, and M. Bennamoun, "A guide to convolutional neural networks for computer vision," *Synthesis Lectures on Computer Vision*, vol. 8, no. 1, pp. 1–207, Feb. 2018, doi: 10.2200/s00822ed1v01y201712cov015.
[26] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.
[27] C. M. Bishop, *Pattern recognition and machine learning*, 1st ed. New York: Springer New York, 2006.
[28] M. Z. Al-Dabagh, M. H. Mohammed Alhabib, and F. H. AL-Mukhtar, "Face recognition system based on kernel discriminant analysis, k-nearest neighbor and support vector machine," *International Journal of Research and Engineering*, vol. 5, no. 2, pp. 335–338, Mar. 2018, doi: 10.21276/ijre.2018.5.3.3.
[29] F. Mezzoudj and A. Benyettou, "On the optimization of multiclass support vector machines dedicated to speech recognition," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7664, no. 2, Springer Berlin Heidelberg, 2012, pp. 1–8.
[30] J. A. Lasserre, C. M. Bishop, and T. P. Minka, "Principled hybrids of generative and discriminative models," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2006, vol. 1, pp. 87–94, doi: 10.1109/CVPR.2006.227.
[31] C. E. Thomaz and G. A. Giraldi, "A new ranking method for principal components analysis and its application to face image analysis," *Image and Vision Computing*, vol. 28, no. 6, pp. 902–913, Jun. 2010, doi: 10.1016/j.imavis.2009.11.005.
[32] A. Kumar and S. Shekhar, "Personal identification using multibiometrics rank-level fusion," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 41, no. 5, pp. 743–752, Sep. 2011, doi: 10.1109/TSMCC.2010.2089516.
[33] P. Ghuli, S. B. N, and A. G. Rao, "Development of framework for detecting smoking scene in video clips," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 13, no. 1, pp. 22–26, Jan. 2019, doi: 10.11591/ijeecs.v13.i1.pp22-26.
[34] B. Ammour, L. Boubchir, T. Bouden, and M. Ramdani, "Face–iris multimodal biometric identification system," *Electronics*, vol. 9, no. 1, Jan. 2020, doi: 10.3390/electronics9010085.
[35] M. Hammad, Y. Liu, and K. Wang, "Multimodal biometric authentication systems using convolution neural network based on different level fusion of ECG and fingerprint," *IEEE Access*, vol. 7, pp. 26527–26542, 2019, doi: 10.1109/ACCESS.2018.2886573.

[36] S. Soleymani, A. Torfi, J. Dawson, and N. M. Nasrabadi, "Generalized bilinear deep convolutional neural networks for multimodal biometric identification," in *Proceedings - International Conference on Image Processing, ICIP*, Oct. 2018, pp. 763–767, doi: 10.1109/ICIP.2018.8451532.
[37] M. Farmanbar and Ö. Toygar, "Feature selection for the fusion of face and palmprint biometrics," *Signal, Image and Video Processing*, vol. 10, no. 5, pp. 951–958, Jul. 2016, doi: 10.1007/s11760-015-0845-6.

BIOGRAPHIES OF AUTHORS

Chahreddine Medjahed computer science graduate from Mohamed Boudiaf University of Science and Technology Oran Algeria, is currently Ph.D. student in computer science. He is an assistant professor at the Faculty of Computer Sciences of Hassiba Benbouali Chlef, Algeria. His research focuses on Machine Learning and Deep Learning methods and their applications in Image and Video Processing. He can be contacted at email: chahreddine.medjahed@univ-sba.dz.

Abdellatif Rahmoun in the past 20 years; I have been intensively involved in AI research like: neural nets, fuzzy systems, genetic algorithms, genetic programming, hybrid intelligent systems. Presently, working in biometrics, grammar-based genetic programming, strongly typed GP. Recently, I focus on new trends in embedded systems, IoT, smart devices, cloud computing and big data. Presently, I am a full professor at the higher school of computer science (ESI-SBA, Algeria). I am also involved in several international academic societies, as a reviewer or editorial board member in some computer science international journals. You can contact me on my email: a.rahmoun@esi-sba.dz.

Christophe Charrier obtained his Ph.D. in Computer Science from the University Jean Monnet (Saint-Etienne), in 1998. In 2008, he joined the GREYC Lab and the e-payment and biometrics research unit. He researches focuses on Image Quality assessment, Human Vision and Biometrics. He can be contacted at email: christophe.charrier@unicaen.fr.

Freha Mezzoudj received her Ph.D. in 2018 from the University of Science and Technology Oran Mohamed Boudiaf (USTO-MB), Algeria, in the field of Artificiel Intelligence (AI) and Pattern Recognition. I also have received a Magister degree in 2010 from the same University USTO-MB, in the field of AI. Actually, I am an Associate Professor at the department of Computer Sciences of the University Hassiba Benbouali Chlef, Algeria. My subjects of researche are in the fields of Natural Language Processing, Speech Recognition and Machine/Deep Learning. I am the author of more than 10 conference papers and publications. You can contact me on my professional email: f.mezzoudj@univ-chlef.dz.